Hipótesis Z

David González

3/5/2022

# Pruebas de hipótesis

## Promedio con Distribución Z para 1 población.

Regresando al mismo problema tomado para un Intervalo de Confianza de un promedio para 1 población tenemos:

**Se quiere analizar el saladio de las personas entre 18 años y 60 años. Una persona afirma que el salario promedio de este grupo es mayor a los 800000 colones.**

Del enunciado tenemos las siguientes hipótesis:

Basándonos en sabemos que el tipo de prueba es de *Cola Derecha* utilizando la distribución .

Para automatizar el proceso se utilizará del paquete stest la función z.test que contiene:

stests::z.test(x, sigma2, mu, alternative, conf.level)

en donde:

| Argumentos | Descripción |
| --- | --- |
| **x:** | vector de valores numéricos |
| **sigma2:** | varianza poblacional conocida |
| **mu:** | media de la hipótesis nula |
| **alternative:** | es el tipo de prueba, se puede escoger entre two.sided para casos de dos colas, greater para caso de cola derecha y less para el caso de cola izquierda. |
| **conf.level:** | nivel de confianza dle intervalo, por defecto viene en 0.95 |

A continuación se mostrará las variables que se requieren para realizar la prueba de hipótesis de una media:

| Variables | Tipo | Descripción |
| --- | --- | --- |
| **salario:** | numérica | salario de personas entre los 18 años y 60 años |
| **salarioVar:** | numérica | varianza muestral de la población |

La implementación en R para las variables es la siguiente:

# Tomamos los salarios de personas entre los 18 años y 60 años de edad  
salario <- Elecciones[Elecciones$edad > 17 & Elecciones$edad < 61,]$salarioMes

#La variable salario la convertimos a tipo numeric.  
salario <- as.numeric(salario)

Obtenemos la varianza:

#Salario varianza  
salarioVar = var(salario)

A continuación la implementación de la prueba de hipótesis:

stests::z.test(x = salario, sigma2 = salarioVar, mu = 800000, alternative = "greater", conf.level = 0.95)

##   
## One Sample z-test  
##   
## data: salario  
## Z = -0.7025, p-value = 0.7588  
## alternative hypothesis: true mean is greater than 8e+05  
## 95 percent confidence interval:  
## 657029.2 Inf  
## sample estimates:  
## mean of salario   
## 757212.9

De lo obtenido podemos observar que la Región de Aceptación *(RA)* y la Región de Rechazo *(RR)* quedaron de la siguiente manera:

Notamos que para este caso no se encontró evidencia suficiente en contra de por lo que no se rechaza, pues:

Llegando a concluir que el salario promedio a lo sumo ronda los 800000 colones.

## Diferencia de promedios con Distribución Z (2 poblaciones).

Tomando como base las poblaciones utilizadas para un Intervalo de Confianza de una diferencia de promedios usando distribución Z, tenemos lo siguiente:

Sean X e Y variables aleatorias, donde:

**X:** Población de personas menores a 40 años.

**Y:** Población de personas mayores a 39 años.

**Nos planteamos si la media del interés en la política es la misma para el grupo de personas menores a 40 años y el grupo de personas mayores a 39 años.**

Del enunciado tenemos las siguientes hipótesis:

El tipo de prueba es de *Dos Colas* por lo que nos presenta. También se utilizará la distribución

Para la implementación se utilizará la función z.test del paquete BSDA:

BSDA::z.test(x, y, sigma.x, sigma.y, mu, alternative, conf.level)

en donde:

| Argumentos | Descripción |
| --- | --- |
| **x:** | vector de valores numéricos correspondiente a la población 1 |
| **y:** | vector de valores numéricos correspondiente a la población 2 |
| **sigma.x:** | desviación estándar de la población 1 |
| **sigma.y:** | desviación estándar de la población 2 |
| **mu:** | diferencia de medias especificada en la hipótesis nula |
| **alternative:** | tipo de prueba, se puede escoger entre “two.sided” para casos de dos colas, “greater” para caso de cola derecha y “less” para el caso de cola izquierda. |
| **conf.level:** | nivel de confianza dle intervalo. Tiene que ser un valor entre 0 y 1 |

Variables requeridas para realizar la prueba:

| Variables | Tipo | Descripción |
| --- | --- | --- |
| **interesPol1:** | numérica | Valor entre 0 y 10 que se refiere al interés en la política de personas menores a 40 años. |
| **interesPol2:** | numérica | Valor entre 0 y 10 que se refiere al interés en la política de personas mayores a 39 años. |
| **intPolDE1:** | numérica | desviación estándar del grupo1 |
| **intPolDE2:** | numérica | desviación estándar del grupo2 |

A continuación se muestra la implementación:

#Seleccionamos a las personas menor de 40 años junto con el interés hacia la política que tienen  
interesPol1 <- Elecciones[Elecciones$edad < 40 ,]$interesPolítica

#Seleccionamos a las personas de 40 años o más, junto con el interés hacia la política que tienen  
interesPol2 <- Elecciones[Elecciones$edad >= 40 ,]$interesPolítica

# Sacamos la desviacion estándar del primer grupo  
intPolDE1 <- sd(interesPol1)  
intPolDE1

## [1] 2.599643

#Desviacion estándar del segundo grupo.  
intPolDE2 <- sd(intPolDE2)  
intPolDE2

## [1] 2.822799

A continuación la implementación de la prueba de hipótesis

BSDA::z.test(x = interesPol1, y = interesPol2, alternative = "two.sided" , mu = 0, sigma.x = intPolDE1, sigma.y = intPolDE2 ,conf.level = 0.90)

##   
## Two-sample z-Test  
##   
## data: interesPol1 and interesPol2  
## z = -0.32578, p-value = 0.7446  
## alternative hypothesis: true difference in means is not equal to 0  
## 90 percent confidence interval:  
## -0.6320728 0.4230857  
## sample estimates:  
## mean of x mean of y   
## 5.237200 5.341694

De lo obtenido podemos observar lo siguiente:

Analizando los resultados, se concluye que no hay suficiente evidencia para rechazar ya que el valor Por lo tanto lo planteado en el problema se concluye que es altamente probable que las medias de ambas poblaciones sean iguales.